THE EDUCATION OF EXAMINERS.[1]

THE subject that I have chosen for my presidential address may at first sight seem far from inviting. Yet, in spite of the unusual title of my paper, I undertake to say that most of you present here to-day will follow the results which I shall lay before you with ease, and will find a growing interest in certain ideas which cannot but prove novel to those of you who have not before thought of examiners as belonging to the human race, and therefore capable of education.

In a sense we are all examiners. We note and tabulate events and their causes. We distribute mankind into ethnological groups, or compare them as industrial workers. We ascertain their wants and their means of satisfying those wants. We examine and record the growth of custom, the physical and mental development of the human being, the changes in the mind itself and the order of such changes, the progress and decay of language, the distribution of wealth, the progress of society. Even the laws of statistics are submitted to examination.

Thus, side by side with the advance of theory in connection with all the sciences that fall under this section (archæology, education, mental science, philology, political economy, sociology, statistics), goes the scrutiny of results. It is justifiable, therefore, to think that an examination of methods of examination, even in connection with only one of those subjects, will throw a light upon such methods in general. I propose to-day to consider that small part of education which consists in the testing of the results of study by written papers.

You will perhaps wonder how it is that I have taken such an interest in the doings of examiners. The fact is that I am one of the few persons who have been for a lengthy period in the position of an examiner of examiners. In the position which I held in the Civil Service Commission for nearly fifteen years, it was my daily task to consider the character of the papers set by some of the highest dignitaries at Oxford and Cambridge, and other universities, to candidates for appointments in the English Civil Service. I had, moreover, to investigate the marking of the written answers of candidates, and to say whether the general results appeared to me to be fair and trustworthy.

Of course, it will be understood that there are good as well as bad examiners. If the methods of good examiners are compared together, it will be found that they tend to uniformity, and that their results have certain characteristics in common. Whereas the methods and results of bad examiners differ from one another in every conceivable way.

But how are these results to be shown? It is not possible to obtain such information by running the eye down the totals awarded to candidates in the mark-sheets. Patient study will no doubt do something, but, where figures occur irregularly, it is hard to appreciate their import without definite classification.

In these days of the almost universal use of "squared" paper, all that is required is to find the percentages of candidates obtaining marks between the limits named, and to mark them off by counting the squares, say five candidates to a square. If the maximum in the subject is not 100, then it is only necessary to reduce the marks to that scale. By joining the top points of the vertical lines, which we call ordinates, the characteristic curve of the examiner is obtained, or, what is even more satisfactory, if black columns are raised on the bases 0 to 10, 11 to 20, &c., to show the number of candidates within these limits of marks, the result is a number of stepping-stones, shown in silhouette, and rising and falling in general harmony with the curve.

Difficulties presented themselves to me as soon as I began to plot the results of examiners from their mark-sheets. Until this had been done it was impossible to analyse the character of the marking, even after hours of study of the mark-sheets themselves. But as soon as the graphical representation had been arrived at, the whole matter was simplified. It was only necessary to determine whether there was any special form of curve to which the many varieties that have been placed before you ought to tend, or whether each subject, and even each examiner, might be properly represented by a different curve. I very soon became convinced that there was a tendency among the best examiners in many subjects to obtain results which gave the graphical form of a gendarme's hat (Fig. 2).

This form is one which is recognised by mathematicians as belonging to the so-called curve of "errors." I can best illustrate what is meant by this curve by supposing that some person in this room, experienced in the use of fire-arms, were asked to fire shots at a paper target on which a vertical straight line had been drawn as the mark to be aimed at. After a large number of shots had been fired, you would find that the holes in the target were arranged in about equal numbers on either side of the line, and that very few had actually hit the mark. If the distance of each shot from the centre line were measured and entered on a table, we should find so many falling within one inch of the line, so many between one inch and two inches, and so on. The curve now placed before you (Fig. 1) is produced by showing the number of shots falling within one inch on one side as a column of proportionate height erected on a base reaching one inch from the centre line. Similarly the column showing the number between one and two inches is drawn on a base between one and two inches from the centre line, and so on.

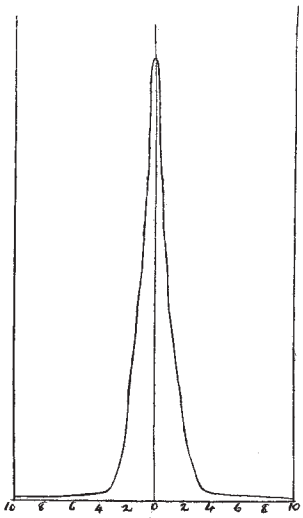


FIG. 1.—Curve showing Pistol Practice at Vertical Line (good shot).

Now I show you a second curve (Fig. 2), in which the pistol has been put into the hands of an inexperienced person. You will at once perceive that these two curves are familiar to you. The curve of the good shot resembles the curve of the bad examiner, and the curve of the bad shot the curve of the good examiner. I think you will spare me giving you the mathematical equation of this curve, although many of the theorems and problems connected with it are extremely interesting. In preparing my paper to-day I have had to consider some of these questions from a mathematical point of view, and in doing so I have had the inestimable assistance of Miss Fawcett. I do not, however, propose to weary you with the mathematical treatment of the subject, but one result deserves consideration, because it is at the root of all the properties of this curve. If we allow the two sets of shots to be fired at one target, and classify them as before (dividing each total by two, since the number of shots is doubled), we shall obtain a curve of the same family as the component curves. However many times the process is repeated, each marksman will repeat his identical curve—on the supposition that he does not improve owing to practice—and of course the resultant curve due to both sets will be repeated.

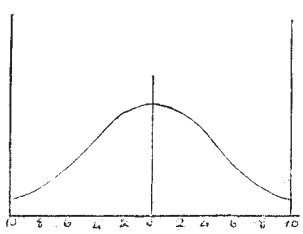


FIG. 2.—Curve showing Pistol Practice at Vertical Line (inexperienced shot).

Instead of taking only two performers with the pistol of unequal merit, we may bring within our view a considerable number in an ascending or descending scale of accuracy, and trace upon one sheet a series of these curves. Here is such a series (Fig. 3).

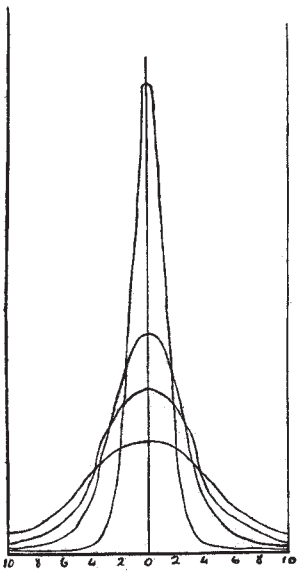
In each of these curves it should be noticed that the extreme portions never touch the base line, but they approach closer and closer to that line, so that the area enclosed in each case between it and the curve in question

[1] Abridged from an address delivered before Section D of the South African Association for the Advancement of Science on April 5 by Mr. E. B. Sargant, Education Adviser to Lord Milner.

depends upon a portion on each side of the middle ordinate which is at a measurable distance from that ordinate. Where the practice is accurate, the portion of the whole figure that may be safely excluded in calculating the area is much larger than in the cases where the shooting is wild.

A measure of the accuracy of the marksmen is obtained by drawing an ordinate to divide into equal parts the half area to the right or left of the middle ordinate, and estimating the distance between these two ordinates.

The whole area under consideration represents the total number of shots, and is therefore the same in the case of each curve. For the sake of simplicity we may suppose that 100 shots are fired. It is not true that that number of shots will in any case give the exact curve. We should only obtain its precise form by firing an infinite number of shots and then reducing the whole to a percentage. But for the sake of simplicity in our argument we will talk of 100 shots as the number that has been fired, and say that the area is proportional to that number. We see, then, that all the areas enclosed by each of these curves respectively and the base line are equal; and this gives us a simple way of plotting any one series if a single curve has been drawn. It is only necessary to suppose the curve to be stretched to a certain extent in either the horizontal or vertical direction, and to be contracted to a proportionate extent in the other direction, in order to pass to another curve of the series. In fact, if one of the curves were painted on a stretched india-rubber sheet, all the other curves could be got from it by pulling the sheet in one direction and slacking it off in the other.



Deviation in inches.

FIG. 3.—Series of Curves of "Error."

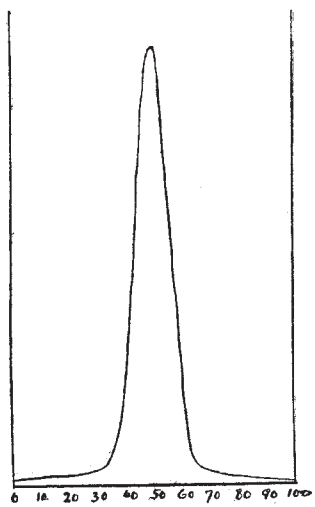
Another plan would be to bend a loop of wire into the form of one of the curves, and to place a lamp behind it so as to throw the shadow upon a screen. The loop and lamp might then be easily made to move in such a manner that the shadows in the successive positions gave the whole series of curves.

You will notice in the figure the points which show the intersection of neighbouring curves with one another. This is called, in mathematical language, the envelope of the family of curves. In this case it is a portion of two rectangular hyperbolæ.

Now, instead of our performers with the pistol, let us take the case of a series of examiners. As soon as I had observed that the curves of good examiners tended to approximate to the curve of errors, I cast about for the reason of this similarity. It is not far to seek. If we consider one particular candidate as the mean candidate, that is, a candidate such that there are as many above him as below him, we shall see how natural it is that the candidates should group themselves about this central figure as the pistol shots about the mean shot. It is clear that the curve of the good examiner should resemble the curve of the bad shot. The object of examination is to separate the candidates from one another as widely as is permissible under the given conditions, while the object of the target-practice is to get as many shots near the central line as possible.

Here we come to a most important limitation. You have already noticed that the curves we have been considering never touch the base line, that is to say, given a sufficient number of candidates, there will always be one or two removed to an extraordinary degree from the bulk of their fellows. But the examiner is obliged to give marks within certain limits, which he fixes arbitrarily as 0 and 100. If he were to place his zero point at a very great distance from the middle point, representing 50 marks, he would be able, no doubt, to make allowance for extraordinary candidates; on the other hand, the bulk of the candidates would be placed so close together that he would not be able to distinguish between them in any satisfactory manner. He is therefore bound to choose points such that the areas enclosed between the base line and the curve which lies beyond those points are very small compared with the areas up to the middle line. All the candidates beyond those points must be considered as having either nought or full marks.

Now you will see, I think, how an examiner in English composition, especially if he is a university man who has become acquainted with the finest examples of literature, tends to get a very steep form of curve (Fig. 4). He looks at some one paper, which differs to a considerable extent as regards both style and matter from the mean paper, and says, "This paper should have 80 marks at least." But then he thinks, perhaps unconsciously, "How do I know that, before finishing the pile of papers before me, I shall not find a budding Milton or Addison or Charles Lamb? If I give this candidate 80 marks, shall I be able to assign its true value to a composition of such extraordinary merit?" So he only awards 60 marks to the composition, and finds almost certainly, when he comes to the end of his pile, that no candidate has received any mark near 100. It is too late now to begin marking the papers all over again, and accordingly he sends in returns which do not serve to distinguish between the candidates in English composition to the same extent as they are distinguished in geometrical drawing, for example. The result is that a good candidate in the former subject is treated unfairly as compared with a good candidate in the latter subject.



Number of marks gained.

FIG. 4.—English Composition.

Again, we see why a curve (Fig. 5) based on marking dictation papers by the system of deductions is so abnormal. In this case, the examiner, without considering minor defects, makes a certain deduction for each mistake in spelling. If 10 marks are taken off for each mistake, all candidates having more than ten errors receive no marks, whereas if we were to assign negative marks, the curve of errors would almost certainly be reproduced, the mean ordinate being below the zero point. The divergence which you perceive near the point representing full marks is due to there being a good many candidates who make no important mistakes. If minor defects, such as refinements of punctuation, were considered, and the scale stretched beyond 100, this divergence would also disappear.

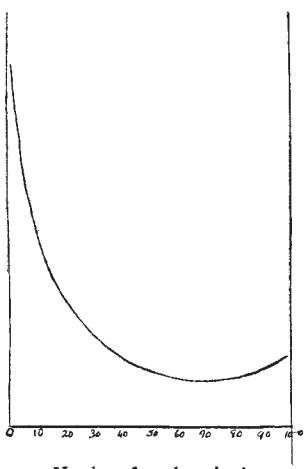
The problem which presented itself was how to bring these very different results into some accord. In order to give equal weight to various subjects having the same maximum, it seemed to me necessary that the examiners should have a common standard to work up to. Accordingly, during the latter period of my service with the Civil Service Commission, I caused such a diagram as has been placed before you to be printed on the sheet containing the examiner's report of his work. On that diagram, also, was printed a curve resembling a moderate sized gendarme's hat. If, as often happened, the examiner had 1000 papers to mark, he was requested to go through a batch of 100 taken at hazard, and to plot his curve upon the diagram. After a few examinations an old hand would probably find that his curve for the first 100 resembled closely the standard curve before him, but a fresh examiner might find himself altogether beside the mark. In such a case he was asked

to put aside the first 100 papers and to begin marking the fresh papers on such different lines as would, in his judgment, produce an approximation to the normal curve. On the supposition that he had achieved that result for the second 100, and continued to find that his curve was pretty constant for the third 100, fourth 100, and so on, he was asked at the end of all the papers to re-mark the first 100.

You might imagine that many examiners disliked having to place themselves upon this bed of Procrustes, but in the generality of cases it was not so. They positively took a delight in examining themselves. The process became one of self-education in marking.

Before leaving this part of my subject I should like to warn you that certain causes, which an examiner cannot always control, may make it difficult to obtain such an ideal curve as I have shown. It is not possible for me to enter fully into this part of the subject, but I will point out one cause at least that he can control—I mean the examination paper.

Good marking will not compensate for a bad paper. Every candidate must have his chance, in some question or other. Otherwise the examination is like a hurdle-race in which the hurdles are so high that a considerable number of candidates find themselves stopped from reaching the goal at all. The curve, in such a case, tends to assume a shape of this kind, mounting very rapidly to the zero line (Fig. 5)—just the curve, in fact, which we have already seen in connection with a dictation paper. In this case it is not the marking which is wrong, but the examination paper.



Number of marks gained.

FIG. 5.—Dictation Paper.

Accordingly, I found in practice that it was necessary to point out to examiners, before ever their papers were proposed in manuscript, that they ought to divide their questions roughly into (say) three portions, of which one portion could be answered by candidates of inferior power, a second should be within the range of mediocre candidates, and a third only possible to candidates who might be classed as good to excellent. The result of these directions was that examiners soon found little difficulty in spreading out their candidates in the desired way. In setting their questions they had before their eyes the little gendarme's hat.

Among the causes, beyond the control of the examiner, which may interfere with the formation of this curve, we must reckon as in the first rank:—(1) such a small number of candidates as does not give fair play to the law of probabilities; (2) any selection of candidates by a preliminary examination or other means.

With regard to the causes just named, I will only say that it has been found that the method can be applied successfully when there are not less than one hundred candidates, and that, even below this number, the curve, though irregular in formation, gives us very useful information as to an examiner's capacities. With regard to the second cause, a great deal can be done to produce a satisfactory curve by setting such questions in the further papers as are only addressed to the candidates who remain after the preliminary sifting.

I trust that I have now fulfilled the promise with which I started, namely, to show you how examiners themselves may be examined; and not only this, but you will understand that it is possible to educate examiners so as to enable them to form a much more accurate and sustained judgment of a large number of candidates than would have been within their power without such preliminary guidance.

THE CHEMICAL REGULATION OF THE SECRETORY PROCESS.[1]

THE researches which we wish to bring briefly before the Royal Society deal with the mechanism of adaptation to changes in the food and the chemical correlation of the activities of different organs engaged in the digestion and assimilation of the food.

According to Pawlow, the secretion of the pancreatic juice is exactly comparable to the secretion of saliva, and is effected by a nervous reflex. The starting point of this reflex is the stimulation of the duodenal mucous membrane by the chyme, or by substances such as oil, ether, or oil of mustard. Not only is the pancreatic juice turned out into the intestine just at the time when it is required, but, according to Pawlow, the composition of the juice varies according to the food, the proteolytic ferment being increased by a diet of meat, while the amylolytic ferment is increased by a starchy diet. This adaptation of the glandular activity was ascribed by him to a species of "taste" in the mucous membrane. It was imagined that the different constituents of the food excited different nerve endings, which, in their turn, caused reflex activity of different mechanisms in the pancreas itself. The field of these assumed reflexes was considerably narrowed by the researches of Popielski (*Gazette Clinique de Botkin*, 1900) and Wertheimer (*Journal de Physiologie*, vol. iii. p. 335, 1901), who showed that the introduction of acid into the duodenum was productive of secretion even after destruction of all nerve connections of the pancreas and alimentary canal with the central nervous system, and even after extirpation of the sympathetic ganglia of the solar plexus. It was with a view to determine the mechanisms of this reflex secretion of the pancreas, as well as of the adaptation of the pancreatic secretion to variations in the food of the animal, that we began our researches.

The last named authors had also shown that the secretion occurred, but in smaller quantities, if the acid was inserted in any part of the small intestine, with the exception of the lower end of the ileum. It was thus easy to examine the effects of the introduction of acid into a loop of ileum in which all nerve connections with the pancreas, or with the rest of the body, had been destroyed. This crucial experiment had, curiously, not been performed by previous workers in the subject. On carrying it out, we found that destruction of all nerve connections made no difference to the result of introducing the acid. The pancreatic secretion occurred as in a normal animal. It was therefore evident that we had to do here with a chemical rather than a nervous mechanism. Previous work had narrowed the question down to such a degree that the further steps were obvious. We knew already that the introduction of acid into the blood-stream had no influence on the pancreas; hence the acid introduced into the intestine must be changed in its passage to the blood-vessels through the epithelial cells, or must produce in these cells some substance which, on access to the blood stream, evoked in the pancreas a secretion. This was found to be the case. On rubbing up the mucous membrane with acid, and injecting the mixture into the blood-stream, a copious secretion of pancreatic juice was produced. It was then found that the active substance, which we call *secretin*, was produced by the action of acid from a precursor in the mucous membrane, probably in the epithelial cells themselves. Once formed by the action of acid, it could be boiled, neutralised, or made alkaline, without undergoing destruction. The precursor of the substance (*pro-secretin*) cannot be extracted by any means that we have tried from the mucous membrane. Even after coagulation of the mucous membrane by heat or alcohol, however, secretin can still be extracted from the coagulated mass by the action of warm dilute acid.

We have not yet succeeded in determining the chemical nature of secretin, though we have obtained chemical evidence which will serve to exclude certain classes of substances. Thus the fact that it will stand boiling shows that it is neither a coagulable proteid nor a ferment. It is soluble in 90 per cent. alcohol in the presence of ether, but it is insoluble in absolute alcohol and ether. It is slightly diffusible through animal membranes. It can be

[1] Abstract of the Croonian Lecture. By Dr. W. M. Bayliss, F.R.S., and Prof. E. H. Starling, F.R.S. Read before the Royal Society, March 24.